Structure, Vol. 13, 1047–1054, July, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.04.009

Scaling Behavior and Structure of Denatured Proteins

Feng Ding,* Ramesh K. Jha, and Nikolay V. Dokholyan
Department of Biochemistry and Biophysics
School of Medicine
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599

Summary

An ensemble of random-coil conformations with no persistent structures has long been accepted as the classical model of denatured proteins due to its consistency with the experimentally determined scaling of protein sizes. However, recent NMR spectroscopy studies on proteins at high chemical denaturant concentrations suggest the presence of significant amounts of native-like structures, in contrast to the classical random-coil picture. To reconcile these seemingly controversial observations, we examine thermally denatured states of experimentally characterized proteins by using molecular dynamics simulations. For all studied proteins, we find that denatured states indeed have strong local conformational bias toward native states while a random-coil power law scaling of protein sizes is preserved. In addition, we explain why experimentally determined size of the protein creatine kinase does not follow general scaling. In simulations, we observe that this protein exhibits a stable intermediate state, the size of which is consistent with the reported experimental observation.

Introduction

Unveiling the structural and dynamic properties of denatured proteins is crucial for understanding the protein folding problem. For example, computational determination of protein thermodynamic stability and NMR hydrogen exchange experiments rely on models of the protein unfolded states. It has long been postulated that the denatured state of proteins is comprised of an ensemble of featureless random-coil-like conformations. According to Flory's theory (Flory, 1965), the size of random-coil polymer chains, the radius of gyration, Rg, follows a power law dependence on the length of the polymer chain n, $Rg = R_0 n^{\nu}$. Here, R_0 is the scaling constant, which is the function of persistent length, and ν represents the power law scaling exponent. Flory predicted the exponent to be 0.6, and, later, a more accurate renormalization calculation obtained $\nu = 0.588$ (Le Guillou and Zinn-Justin, 1977). Tanford et al. (1967) first confirmed this random-coil scaling behavior for denatured proteins. Using intrinsic viscosity measurements for 12 proteins denatured by 5–6 M GuHCl, the authors obtained a scaling exponent of $\nu = 0.67 \pm 0.09$. Wilkins et al. (1999) have shown that the hydrodynamic radii of sets of eight highly denatured, disulfide-free proteins fit power law relationships with $\nu = 0.58 \pm 0.11$. Recently, Kohn et al. (2004) reassessed the scaling behavior of denatured proteins by using small-angle X-ray scattering (SAXS) for 17 proteins of lengths varying from 8 residues to 549 residues. The Rg values for 11 other proteins were taken from literature. It was shown that 26 out of 28 exhibit a power law scaling with the exponent $\nu = 0.598 \pm 0.029$. All of these experimental results confirm the random-coil scaling of denatured proteins. In a random-coil model of denatured state, the proteins are believed to lack persistent structures both locally and globally. The distribution of end-to-end distances or radii of gyration can be well fitted by a Gaussian distribution. A recent computational study (Goldenberg, 2003) confirmed this behavior by generating an ensemble of protein conformations whereby only steric interactions between amino acids were considered for four different proteins. A scaling component of 0.58 ± 0.02 was obtained.

The random-coil scaling of protein sizes (Flory, 1965; de Gennes, 1979) was originally derived for polymers in which specific interactions were not considered. However, proteins are heteropolymers for which specific interactions between amino acids play an important role and determine a unique native structure. Here, we argue that while the scaling of the sizes of denatured proteins follows the random-coil scaling, as shown in experiments, it does not necessarily mean that the denatured proteins lack native-like structures. Under denaturing conditions, the specific interactions can still play the role of native conformational bias and retain a certain amount of residual native structures. Mounting experimental evidence (Shortle and Ackerman, 2001; Bu et al., 2001; Yi et al., 2000) supports residual native-like topologies in the denatured states for a variety of proteins. Using residual dipolar couplings (RDC) from NMR measurement, Shortle and Ackerman (2001) showed that native-like topology persists under strong denaturing conditions as high as 8 M urea for a truncated staphylococcal nuclease. By applying the quasielastic neutron-scattering (QENS) technique on α-lactalbumin, Bu et al. (2001) demonstrated residual helical structure and tertiary-like interactions even in the absence of disulfide bonds under highly denaturing conditions. Similarly, by using triple-resonance NMR, native-like topology has also been observed in protein L (Yi et al., 2000).

Several theoretical and computational studies (Zwanzig et al., 1992; Pappu et al., 2000; Srinivasan and Rose, 2002; Fitzkee and Rose, 2004b) have been reported to address the role of specific interactions in conformational biasing toward the native state in the denatured states. Zwanzig et al. (1992) employed a mathematical analysis of a simple model and showed that a small and physically reasonable energy biases against locally unfavorable configurations. Using a simple force field with only steric and hydrogen bond interactions, Pappu et al. (2000) demonstrated strong native structure preferences for proteins in denatured states. It has been suggested (Zwanzig et al., 1992; Pappu et al., 2000; Srinivasan and Rose, 2002; Fitzkee and Rose, 2004b)

*Correspondence: fding@unc.edu

that the conformational bias of native structures in the denatured state is a possible explanation of Levinthal's paradox (Levinthal, 1968).

Here, we aim to reconcile the seemingly controversial properties of denatured proteins, i.e., the random-coil scaling of their sizes and the presence of residual native structures. Inspired by the recent work of Fitzkee and Rose (2004a), who reproduced the random-coil scaling exponent by using a denatured protein model with fixed secondary structure elements, we examine denatured states of experimentally studied proteins by using molecular dynamics simulations. We perform discrete molecular dynamics (Zhou and Karplus, 1997; Dokholyan et al., 1998) (DMD) simulation studies of thermally denatured states for 18 experimentally characterized proteins. To assure the protein-like folding properties, we apply a commonly used structure-based potential model, the Gō model (Go, 1983), for amino acid interactions. It has been shown that the Gō model can capture folding kinetics (Ding et al., 2002; Borreguero et al., 2004; Dokholyan et al., 2003) and thermodynamics (Borreguero et al., 2002; Clementi et al., 2000; Zhou and Karplus, 1997, 1999). We study the scaling behavior of sizes as well as structural properties of denatured proteins induced by high temperatures. For all of the studied proteins, we find that the denatured states indeed have strong local conformational bias toward native states, while the random-coil scaling is preserved. Our molecular dynamics study provides a conceptual reconciliation between residual native structures and the random-coil scaling exponent of the denatured proteins (Fitzkee and Rose, 2004a). In addition, we find that creatine kinase (Protein Data Bank [PDB] code: 1QK1 [Berman et al., 2000]), which was observed to derivate from the random-coil scaling in experiments (Kohn et al., 2004), has the existence of a stable intermediate state, consistent with the experimentally observed state in sizes. We postulate that the deviation of this protein from scaling is due to the fact that it was not fully denatured in experiments. Indeed, if we further denature this protein in computer simulations, the size of this protein in the fully unfolded state follows the random-coil scaling.

Results

Average *Rg* Values of Denatured Proteins Do Not Change above T_m in Simulations

We first determine the T_m values for all proteins under study by using a protein interaction model in which the Gō parameter is $r = 1$ (see Experimental Procedures). In Figure 1A, we present the dependence of *Rg* on temperature, *T*, in a slow unfolding simulation for protein 1FU6. There is a sharp transition of *Rg* at $T = 0.94$, where the mean *Rg* changes from 14 Å to approximately 35 Å. Thus, the transition temperature for 1FU6 is estimated to be $T_m = 0.94$. Using the same protocol, we estimate T_m values for different proteins and list them in Table 1. Once the protein is denatured after T_m (Figure 1A), the radius of gyration has large fluctuations, while the average value remains stable as the temperature increases. This behavior is in agreement with the experimental observation that *Rg* values do not change significantly with increases of denaturant concentrations after the denaturation midpoint (Kohn et al., 2004). To test whether there are changes in the distributions of *Rg* values at temperatures higher than T_m due to insufficient sampling in the unfolding simulations, we perform equilibrium simulations at two high temperatures, $T = 1.1T_m$ and $T = 2.0T_m$, and study the distributions of the *Rg* values (Figure 1B). We find that the *Rg* distributions at these two temperatures are very similar; the higher temperature simulations sample protein conformations with slightly larger *Rg* values. However, the average *Rg* values computed from the distributions at the two temperatures, 34.83 Å ($1.1T_m$) and 36.15 Å ($2.0T_m$), are within the experimental error bars. Therefore, for the sake of consistency and to ensure complete protein denaturation, we use *Rg* values calculated from equilibrium simulations at $T = 2.0T_m$ in the following discussion.

Computationally Determined Power Law Scaling Exponent of Denatured Protein Sizes Agrees with Experiments

We study the scaling behavior of the *Rg* of denatured states. We set the interaction parameter at $r = 1.0$. We find that the *Rg* values of denatured proteins indeed follow a power law dependence on the protein length *n*; the points in the double-logarithmic *Rg*(*n*) plot are aligned linearly (Figure 2A). A power law fit yields an exponent, $\nu = 0.60 \pm 0.01$. To compare with experiments, we also present the experimental *Rg* values in the same plot (Figure 2A). As discussed by Kohn et al. (2004), the experimental *Rg* values versus protein length, *n*, follow a power law scaling after excluding two proteins: creatine kinase (PDB code: 1QK1) and a short 8 residue peptide, 1N9V. The best fit of the experimental data yields an exponent of $\nu = 0.598 \pm 0.028$ (Kohn et al., 2004). Thus, the exponent derived from our simulations agrees with the experimental one (Kohn et al., 2004). In contrast to experiments (Kohn et al., 2004), in our simulations, the *Rg* for creatine kinase follows the power law scaling. This discrepancy with experiments is discussed below.

The Computationally and Experimentally Determined Protein Sizes Coincide When the Interaction Parameter $r = 0.2$

Although the scaling exponents from simulations and experiments are in agreement with each other, we find that the *Rg* values from DMD simulations with $r = 1$ are, in general, larger than those in experiments (Figure 2A). We hypothesize that the repulsion of the nonnative contacts is responsible for the larger dimensions of denatured proteins. In the unfolded states, the nonnative interactions are most likely to form because the total number of residue pairs is $\sim n^2$, while the number of native pairs is only $\sim n$ ($<<n^2$). A strong repulsion imposed on nonnative contacts in computer simulations results in a large effective excluded volume of each amino acid, which increases the dimension of denatured proteins. To test this hypothesis, we study the dimensions of homopolymers with $r = -1$, an extreme case in which a uniform attraction is assigned for all amino acid pairs. We find that homopolymers have a

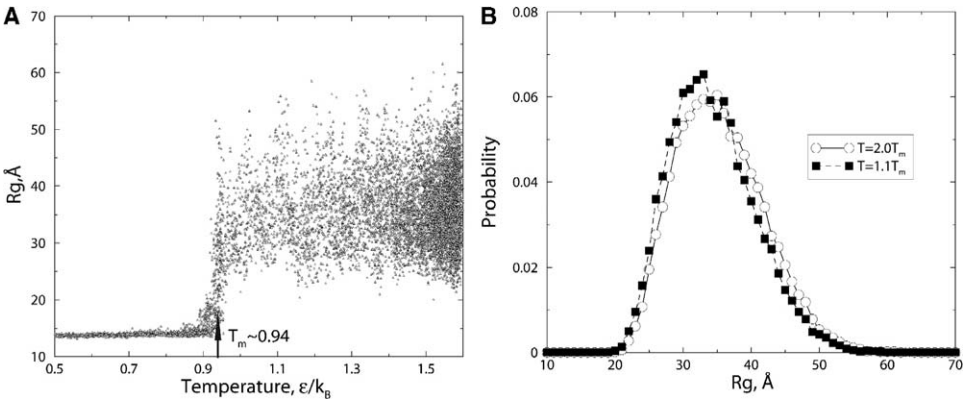

Figure 1. The Computed *Rg* from Simulations

(A) *Rg* as the function of temperature in an unfolding simulation of 1FU6. The simulation is started from folded 1FU6 with a low *Rg* of 14 Å. The protein starts to unfold at T_m = 0.94, with a sharp change of *Rg* to 35 Å. Once unfolded, the *Rg* has large fluctuations, but the average value is approximately stable.

(B) The distributions of *Rg* at two temperatures higher than T_m: 1.1T_m and 2.0T_m. The distributions of *Rg* at these two temperatures are very similar, and the average *Rg* values are within experimental error.

similar scaling exponent, but with smaller *Rg* values than experimental data (Figure 2A). To understand the balance of the native and nonnative interactions that is responsible for the experimentally observed scaling of *Rg* versus protein lengths, we study the effect of the interaction parameter, *r*, on the dimensions of denatured proteins.

We perform equilibrium DMD simulations with *r* values ranging from 0 to 1 and with an increment in *r* values of 0.1. The transition temperatures for different *r* values ranging from 0 to 1 do not change significantly (data not shown). Therefore, we assume the same T_m for different *r* values as the T_m for *r* = 1. To ensure that the proteins fully unfold, we calculate the average *Rg* values from equilibrium simulations at $T = 2.0T_m$. For each value of *r*, we compare simulations and experimental *Rg* values by calculating the root mean square deviation (rmsd) of *Rg* from corresponding experimental values for all studied proteins,

$$\sqrt{\sum_i (Rg_i^{Sim}(r) - Rg_i^{Exp})^2/N}.$$

Here, the subscript *i* corresponds to different proteins under study, and *N* is the number of proteins. Since lengths of some proteins are often different between experiments and available PDB structures (see Table 1), we use the experimental interpolations of *Rg* values for these proteins. Thus, for each value of *r* the rmsd of *Rg* is

Table 1. The Selected Proteins under Simulation Study

Protein	PDB ID	Chain length, *n*		Experimental *Rg*, Å	T_m, ϵ/k_B	Computational *Rg*, Å ($T = 2.0T_m$; r = 0.2)
		PDB	EXP			
FBP-WW domain	1EOL	37	—	—	0.72	17.02 ± 2.71
PKC Cys2 domain	1PTQ	50	—	—	0.85	19.90 ± 3.42
Protein G	2GB1	56	52	23 ± 1	0.87	21.65 ± 3.62
Fyn SH3	1SHF	59	78	25.7 ± 0.5	0.88	23.32 ± 4.02
Protein L (with His tag)	1HZ5	72	79	26.0 ± 0.6	0.93	25.62 ± 4.56
Ubiquitin	1UBQ	76	76	25.2 ± 0.2	0.87	25.73 ± 4.37
pI3K SH3	1H9O	103	103	30.9 ± 0.3	0.94	31.01 ± 5.62
pI3K SH2	1FU6	111	112	29.6 ± 3.3	0.94	32.88 ± 5.97
Rnase A, reduced	1XPT	124	124	33.2 ± 1.0	0.91	36.57 ± 6.68
CheY	1EHC	128	129	38.0 ± 1.0	1.01	34.91 ± 6.18
Lysozyme, reduced	1HEL	129	129	35.8 ± 0.5	0.99	34.50 ± 6.26
Snase	2SNS	141	149	37.2 ± 1.2	1.06	37.82 ± 7.26
Oxymyoglobin	1MBO	153	154	40 ± 2	0.99	37.33 ± 6.76
E. coli DHFR	1DRB	159	167	44 ± 2	1.05	39.95 ± 7.51
C. ablicans DHFR	1AI9	192	—	—	1.07	45.05 ± 7.92
Creatine kinase	1QK1	379	380	46.0 ± 1.5	1.18	66.68 ± 12.67
yPGK	3PGK	416	416	71 ± 1	1.35	68.53 ± 12.28
GroEL	1GRL	518	549	82 ± 4	1.5	76.61 ± 15.68

1EOL, 1PTQ, and 1AI9 are chosen to cover a larger range of protein lengths. Only the *Rg* values for the DMD simulations with r = 0.2 and $T = 2.0T_m$ are shown. The standard deviations of *Rg* from the equilibrium DMD simulations for different proteins are presented.

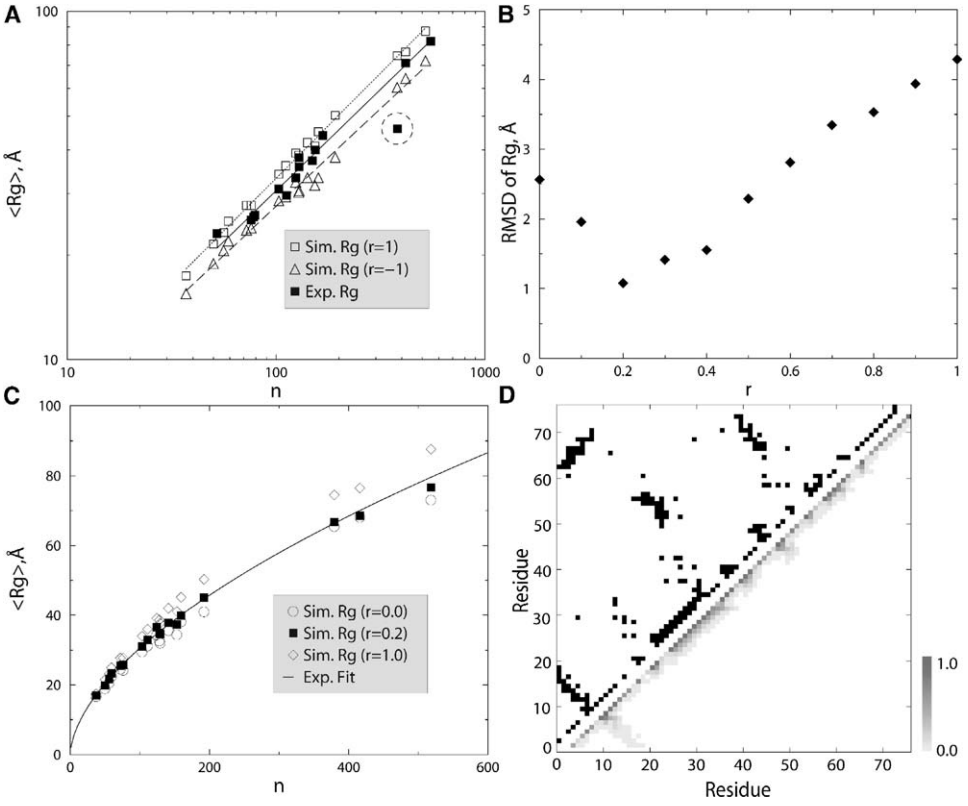

Figure 2. The Scaling and Structure of Denatured Proteins

(A) The log-log plot of the average *Rg* as a function of protein length. Open squares correspond to *Rg*s calculated from DMD simulations using the Gō parameter of $r = 1.0$. The best fit gives $\nu = 0.60 \pm 0.01$ and $R_0 = 2.02$. Open triangles represent the *Rg*s computed from DMD simulations of $r = -1.0$, which correspond to homopolymers. Solid squares correspond to the experimental values.

(B) The rmsd of *Rg* is plotted as the function of *r* by using Equation 1. At $r = 0.2$, the rmsd of *Rg* between computational and experimental values is minimal.

(C) The x-y plot of *Rg* as a function of protein length. The solid line is the experimental best-fit line (Kohn et al., 2004).

(D) The frequency contact map of ubiquitin (1UBQ) for the denatured state ensemble from equilibrium DMD simulations at $2.0T_m$. The upper diagonal is the native contact map; the native contact is presented as a black dot.

$$RMSD = \sqrt{\frac{\sum_i (Rg_i^{Sim}(r,n_i) - R_0^{Exp} n_i^{\nu_{Exp}})^2}{N}}, \qquad (1)$$

where n_i is the length of protein used in simulations and R_0^{Exp} and ν_{Exp} correspond to experimentally determined parameters (Kohn et al., 2004) (Figure 2B). We find that when $r = 0.2$, the rmsd of *Rg* reaches its minimum, and the computationally determined average *Rg* values for $r = 0.2$ agree with the experimental fit (Figure 2C). We also find that the scaling exponents for different *r* values, including $r = -1$, are approximately the same (data not shown). Therefore, a Gō interaction parameter of $r = 0.2$ in the framework of current protein model results in a best fit to the experimental *Rg* values in the denatured state and can be used in future protein studies.

Thermally Denatured Proteins Retain Significant Local Native-like Structures

Our simple protein model with a Gō interaction parameter of $r = 0.2$ reproduces the experimental *Rg* values of all 18 proteins and, thus, the random coil scaling behavior. Next, we study the structural properties of the denatured states. We calculate contact frequencies for all residue pairs in the denatured state generated from DMD simulations. In Figure 2D, we present the contact frequency map for ubiquitin (PDB code: 1UBQ); the contacts are colored according to its frequency. We find that all long-range contacts have frequencies close to zero, while the short-range contacts near the diagonal have large frequencies. For example, the native α helix from residues 20 to 35, the β hairpin from residues 1 to 18, and a short α helix near residue 56 in the denatured state have a high frequency of contact formation in the unfolded state (Figure 2D). In addition, the contact frequency pattern of native β strands retains the specific patterns of contacts between residues *i* and $i + 2$. In contrast, contact frequency maps for unfolded homopolymers have no residual secondary structures at the denatured state, as expected (data not shown). The contact frequency maps of other studied proteins have similar features to ubiquitin.

Creatine Kinase Features a Stable Intermediate, the Size of which Is Consistent with the Experimentally Observed State

As shown in Figure 2A, simulations of 1QK1 result in an approximately twice as large *Rg* than the experimental *Rg* value. However, the computational *Rg* of denatured

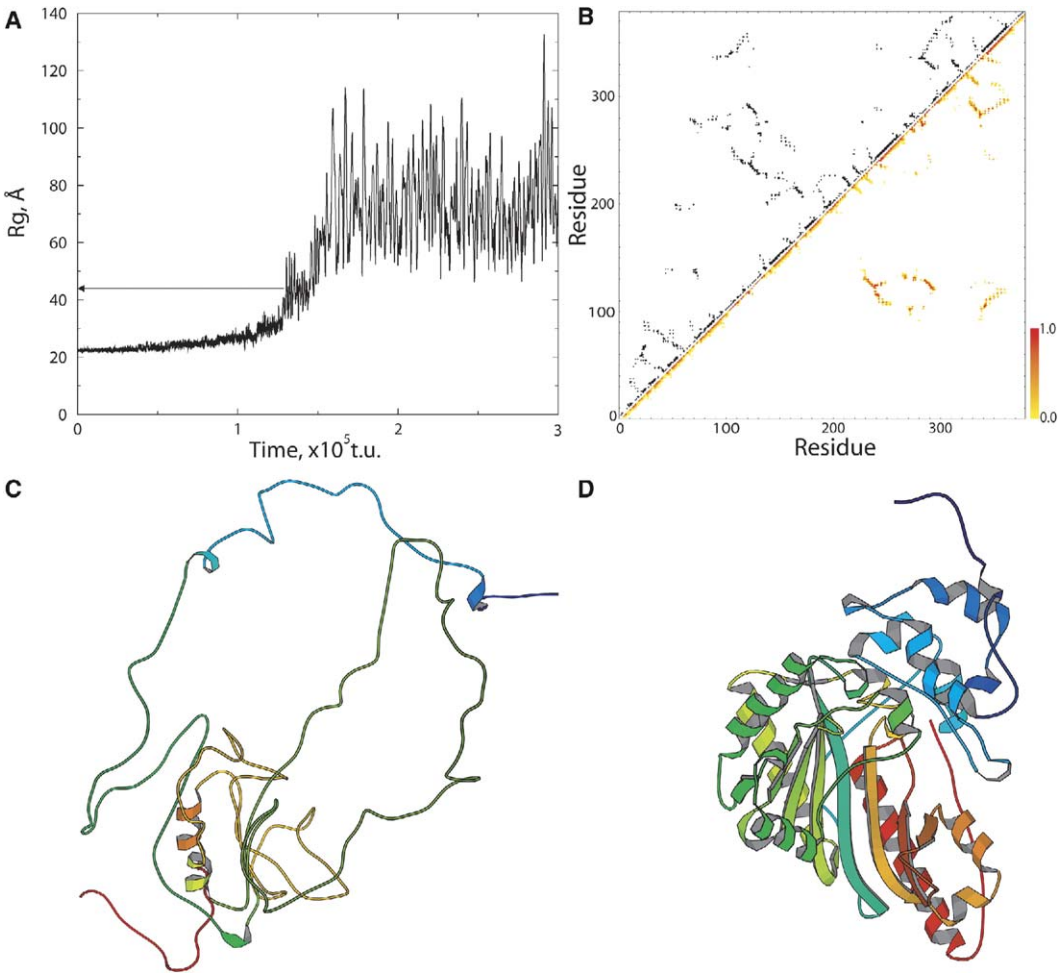

Figure 3. The Intermediate State of 1QK1

(A) The unfolding simulation trajectory of 1QK1 features a folding intermediate state, which has an *Rg* of 43 Å.

(B) The contact frequency map of the 1QK1 intermediate state. Each contact is colored according to its frequency of formation in the intermediate state. The upper diagonal is the native contact map of 1QK1.

(C and D) Cartoon representations of a typical intermediate state and native state of 1QK1, respectively. The proteins are rainbow colored from blue (N terminal) to red (C terminal).

1QK1 follows the power law scaling, while the experimental value is an outlier and is omitted from power law fitting in the work of Kohn et al. (2004). To understand this discrepancy, we examine the unfolding simulation trajectory of 1QK1 with $r = 0.2$ (Figure 3A). Interestingly, we find a stable intermediate state populated along a quasiequilibrium unfolding pathway. The existence of the intermediate state is observed for all the simulations with $r \geq 0$. We compute the average *Rg* of the 1QK1 intermediate state as 43.1 Å, which is close to the experimental value of 46.0 Å (Kohn et al., 2004). To characterize the structural property of the intermediate state of 1QK1, we compute the contact frequency map from the conformation ensemble from the DMD simulation (Figure 3B). We find that the intermediate state of 1QK1 is partially unfolded, with the C-terminal half of the native structure still folded. We present the typical snapshot of the intermediate and native states of 1QK1 in Figures 3C and 3D, respectively.

Discussion

We develop a computational method to model denatured proteins. We use a structure-based Gō potential, a commonly used interaction model in protein folding thermodynamic and kinetics studies (Clementi et al., 2000; Dokholyan et al., 1998, 2000; Yang et al., 2004), to model amino acid interactions. The Gō model enables us to study the scaling behavior and structural properties of denatured proteins by using molecular dynamics simulations. Our simulations suggest that denatured proteins follow the random-coil scaling sizes and retain residual secondary structures akin those observed in the native protein states. Therefore, our model provides a conceptual reconciliation between two seemingly controversial views of protein-unfolded states: while proteins do obey a random-coil scaling, the fluctuations, leading to the formation of local interactions in the unfolded states, are mostly compatible

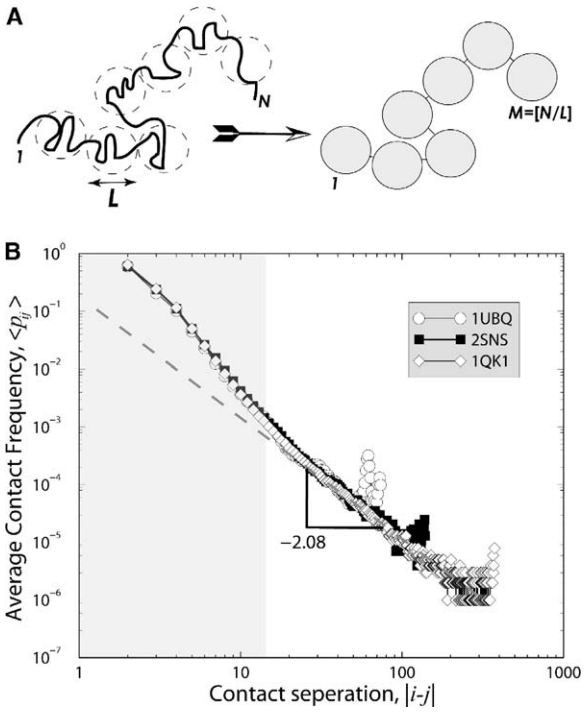

Figure 4. The Physical Origin of the Random Coil Scaling of Denatured Proteins

(A) A scheme for denatured protein chain renormalization. At any instance, a protein of *N* amino acids, which is at thermodynamic equilibrium and is present mostly in the unfolded state, has residual local structures, outlined by the dashed circles (left). The number *L* is the average interaction range along a protein's amino acid sequence, i.e., residue pairs separated by at least *L* amino acids have negligible probability of contact formation in the denatured state. A coarse-graining process groups neighboring *L* residues into a renormalized structure unit, and the denatured protein is reduced to a homopolymer with an effective length of $M = N/L$ (right).

(B) The average contact frequency, $\langle p(|i-j|)\rangle$, as a function of contact length, $|i-j|$, plotted on a double-logarithmic scale. For illustration purposes, we present only plots for three proteins of different lengths: 1UBQ, 2SNS, and 1QK1. The shaded region corresponds to the short-range contacts. The power law fit for the long-range contacts yields a scaling exponent of ≈ -2.08.

with those in native protein states. These local interactions are favored for exactly the same reasons that they are favored in the protein native states: among a number of possible contacts formed in the unfolded state, those contacts that are "natively engineered" to contribute the most to the enthalpy of the protein native states dominate and follow Boltzmann distribution.

In the denatured proteins, what is the physical origin of the random-coil scaling of their sizes along with the seemingly contradictory persistence of local structures? A random-coil scaling exponent, $\nu = 0.6$, can be derived for an excluded volume random walk, originally proposed by Flory (1965). We propose a coarse-graining process, which groups locally interacting amino acids along the polypeptide chain into renormalized structural units (see Figure 4A). This procedure reduces a denatured protein to a renormalized polymer. In order to perform the renormalization, we first calculate the average contact frequency, $\langle p(|i-j|)\rangle$, as a function of contact length, $|i-j|$, for all studied proteins. Here, the average $\langle\ldots\rangle$ is taken over all contacts with the same length $|i-j|$ from simulations at $2T_m$, where the proteins are fully denatured. We find that the $\langle p(|i-j|)\rangle$s for different denatured proteins are strikingly similar (Figure 4B) and depend only on the contact length.

It has been observed that the contact probability, $\langle p(|i-j|)\rangle$, as a function of $|i-j|$ for folded globular proteins is a power law function with the exponent –1.6 ($\langle p[|i-j|]\rangle \sim |i-j|^{-1.6}$) (Dokholyan and Shakhnovich, 2001). In contrast, the contact frequency, $\langle p(|i-j|)\rangle$, as a function of $|i-j|$ for unfolded proteins exhibits two distinct regimes, corresponding to short- and long-range contacts. The scaling of $\langle p(|i-j|)\rangle$ for the long-range contacts is a power law with exponent ≈ -2.0 ($\langle p[|i-j|]\rangle \sim |i-j|^{-2.0}$). This exponent is in agreement with the theoretical prediction of the mean–first passage time of contact formation in polymers (Doi and Edwards, 1988), in which the contact formation is solely due to the chain diffusion. In contrast, $\langle p(|i-j|)\rangle$s for the short-range contacts ($|i-j| \leq 15$ residues; the typical size of secondary elements) are significantly higher than those of the long-range contacts and deviate from the scaling $\langle p(|i-j|)\rangle \sim |i-j|^{-2.0}$ (Figure 4B).

Thus, we define a renormalization length, *L*, "universal" for all proteins, that corresponds to the crossover length of 15 residues. The renormalization process reduces a protein to an effective polymer, which forms contacts solely due to chain diffusion. Protein sizes follow (renormalized) power law scaling as proposed by Flory (1965): $Rg = R_0(N/L)^{\nu} = (R_0 L^{-\nu})N^{\nu}$. Here, the prefactor R_0 is an effective function of the short-range structure in the denatured state, which does not necessarily correspond to the effective Kuhn length (Doi and Edwards, 1988). Thus, we expect that the scaling exponent of denatured proteins, ν, is the same as for homopolymers, whose structural units are locally interacting amino acids.

The observation of residual native secondary structures in the thermally denatured protein states in simulations is consistent with a "guided folding" scenario (Baldwin and Rose, 1999), in which the rate-limiting process of protein folding is the packing of the preformed secondary structures into the correct fold. In contrast, a random-coil model of the denatured state without residual native-like structures implies that a protein has to overcome an excessive entropy barrier to form both the secondary and tertiary structures upon folding. The existence of persistent native-like secondary structures in the denatured state may also be responsible for the recent success of the protein structure prediction using small secondary structure segments derived from the Protein Data Bank (Bradley et al., 2003).

Using DMD simulations and the identified optimal parameter for native and nonnative interactions employed in the DMD potential energy function, we propose an approach to generate conformational ensembles of denatured proteins in agreement with experiments. Here, we use a coarse-grained representation of amino acids that lacks atomic details of unfolded conformations. Further all-atom reconstruction of the unfolded protein conformations can be obtained by adding back the missing backbone and side chain atoms for each amino

acid (F.D. and N.V.D., unpublished data). This generated ensemble of unfolded protein conformations may serve as an alternative to a Gaussian chain model. Our methodology provides an accurate approximation of denatured protein states, which is essential for the calculations of the thermodynamic stability of proteins. In addition, the NMR hydrogen exchange protection factors—a measure of amino acid (backbone amide proton) solvent exposure—depend on the intrinsic exchange rate of the denatured state of proteins, which is usually modeled as a random coil (Bai et al., 1993). Therefore, it is possible to apply the generated ensembles of protein-unfolded states to determine the intrinsic exchange rate of the unfolded states.

Interestingly, our DMD simulations reveal that the scaling exponents for denatured model proteins ($r \geq 0$) and homopolymers ($r = -1$) are all approximately 0.6 (Figure 2). The observed difference between unfolded states of proteins and homopolymers is the scaling constant, R_0, which is a function of persistent length. Our simulations suggest that denatured proteins have a higher probability to form local secondary structures than homopolymers, which leads to a larger persistent length in proteins than in homopolymers. In addition, α-helical versus β strand propensities in proteins alter protein persistent length. Therefore, proteins of the same length but with different secondary structure compositions may have different *Rg* values in their denatured states. The differences might explain the observed fluctuations of *Rg* values in the power law scaling plots of both experiments and simulations (Figure 2A). Further experimental and computational studies of proteins with different secondary structure compositions will help us to understand the scaling behavior of denatured proteins.

We show that the protein 1QK1, the size of which deviates from the random-coil scaling in experiments (Kohn et al., 2004), follows the power law scaling in simulations. The unfolded simulation of 1QK1 suggests an intermediate state with a partially folded subdomain. The size of the intermediate 1QK1 in simulations is approximately equal to the experimentally determined value. We postulate that this folding intermediate state corresponds to the 1QK1 species observed in the experiment. We hypothesize that this intermediate state is resistant to the denaturant GuHCl used in experiments (Kohn et al., 2004) and dominates even at concentration as high as 6 M. Since various types of chemical denaturants have different effects on proteins, we suggest that alternative chemical denaturants or thermal denaturation may be applied on creatine kinase to determine its dimension in the fully unfolded state. Our DMD simulation also predicts the structure of the creatine kinase intermediate state that can be tested in future experimental characterization of this state.

Experimental Procedures

The Coarse-Grained Protein Model

We model proteins as described by Ding et al. (2002). Briefly, we model proteins by beads representing C_α and C_β atoms. In order to mimic the flexibility of real proteins, we apply additional constraints: (1) "covalent" bonds between $C_{\alpha i}$ and $C_{\beta i}$; (2) "peptide" bonds between $C_{\alpha i}$ and $C_{\alpha(i \pm 1)}$; (3) effective bonds between $C_{\beta i}$ and $C_{\alpha(i \pm 1)}$; and (4) effective bonds between $C_{\alpha i}$ and $C_{\alpha(i \pm 2)}$, in which the subscript *i* denotes the amino acid sequence number. The bond parameters, such as lengths and variances, are tabulated by Ding et al. (2002).

We use a Gō-like (Go, 1983) model, similar to one described by Dokholyan et al. (1998), to assign the nonbonded interactions between amino acids. These interactions are determined by the protein native structure. In our model, only C_β atoms that are separated by a distance no larger than 7.5 Å (native contacts) interact via stepwise potentials. The interaction strengths for the native and nonnative contacts are ϵ and $-r\epsilon$ ($r > -1$), respectively. The nearest neighboring C_β atoms experience only hard-core repulsions. Here, *r* is the rescaling parameter for nonnative contact. When $r = -1$, the model reduces to a homopolymer, in which a general attraction exists for any amino acid pairs. In most applications of Gō-like models in protein folding studies (Clementi et al., 2000; Dokholyan et al., 1998, 2000; Yang et al., 2004), *r* is usually 0 or 1. When $r = 0$, the nonnative contacts experience only hard-core repulsion. When $r > 0$, a stepwise repulsion between nonnative contacts is penalized to approach within the cutoff distance. As *r* increases, model proteins have an increasing potential energy gap between the native and nonnative conformations. Zhou and Karplus (1997) have shown that the Gō model features a cooperative folding transitions for $r \geq 0$.

Proteins under Study

From the list of proteins experimentally studied by Kohn et al. (2004), we selected 15 proteins for which PDB structures were available. Our selection was limited by the fact that Kohn et al. used several peptides in the range of 16–39 residues, of which there were no available structures in the PDB. In addition, several other proteins, such as α-TS and carbonic anhydrase, have missing residues or segments in the available PDB structures. In order to improve the coverage range of the protein lengths, we included the FBP-WW domain (PDB code: 1EOL), the PKC Cys2 domain (PDB code: 1PTQ), and DHFR from *Candida ablicans* (PDB code: 1AI9). Since pI3K SH3 (PDB: 1H9O) has 108 residues in the PDB structure, while 103 residues were described by Kohn et al. (2004), we delete 3 residues from the N-terminal side and 2 residues from the C-terminal side that lack any secondary structures. We also notice that there are several proteins with experimentally reported sizes longer than those reported in the PDB. In Table 1, we list the proteins studied here.

DMD Simulations

In DMD simulations, we use a Berendsen (Berendsen et al., 1984) thermostat to control the temperature. The temperature is in units of ϵ/k_B, where k_B is the Boltzmann constant. The proteins are simulated in a cubic box that has dimensions larger than the size of their fully extended conformations. To study the scaling behavior of denatured proteins, it is important to ensure full denaturation of molecules. It has been shown that our protein models can capture the temperature-induced folding/unfolding transition of proteins (Ding et al., 2002). At the folding transition temperature, T_m, model proteins experience sharp transitions in potential energy and *Rg* (Ding et al., 2002).

We estimate proteins' T_m from single quasiequilibrium unfolding simulations. Starting from the native state, we slowly increase the temperature of the system from a native-state favoring low temperature, $T = 0.5$, to a denaturing high temperature, $T = 1.6$. The rate of temperature increase is set to be $10^{-5}\epsilon/k_B$ per time unit (t.u.). We find that for all studied proteins, the unfolding transition occurs within a narrow temperature range. We estimate the T_m from a quasiequilibrium unfolding simulation because we are mainly focused on the unfolded state ensemble at temperatures much higher than T_m. The T_m for a protein is estimated as the temperature at which a sharp transition of *Rg* versus *T* is observed.

Acknowledgments

We thank Sagar Khare and Dr. Kevin W. Plaxco for the insightful discussions and Kyle Wilcox for reading the manuscript. This work is supported in part by Muscular Dystrophy Association grant

MDA3720, Research Grant No. 5-FY03-155 from the March of Dimes Birth Defect Foundation, and the University of North Carolina/International Business Machines Corporation Junior Investigator Award.

Received: January 19, 2005
Revised: March 30, 2005
Accepted: April 3, 2005
Published: July 12, 2005

References

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. Proteins *17*, 75–86.

Baldwin, R.L., and Rose, G.D. (1999). Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem. Sci. *24*, 77–83.

Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular-dynamics with coupling to an external bath. J. Chem. Phys. *81*, 3684–3690.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Borreguero, J.M., Dokholyan, N.V., Buldyrev, S.V., Shakhnovich, E.I., and Stanley, H.E. (2002). Thermodynamics and folding kinetics analysis of the SH3 domain from discrete molecular dynamics. J. Mol. Biol. *318*, 863–876.

Borreguero, J.M., Ding, F., Buldyrev, S.V., Stanley, H.E., and Dokholyan, N.V. (2004). Multiple folding pathways of the SH3 domain. Biophys. J. *87*, 521–533.

Bradley, P., Chivian, D., Meiler, J., Misura, K.M., Rohl, C.A., Schief, W.R., Wedemeyer, W.J., Schueler-Furman, O., Murphy, P., Schonbrun, J., et al. (2003). Rosetta predictions in CASP5: successes, failures, and prospects for complete automation. Proteins *53* (*Suppl 6*), 457–468.

Bu, Z., Cook, J., and Callaway, D.J. (2001). Dynamic regimes and correlated structural dynamics in native and denatured alpha-lactalbumin. J. Mol. Biol. *312*, 865–873.

Clementi, C., Nymeyer, H., and Onuchic, J.N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J. Mol. Biol. *298*, 937–953.

de Gennes, P.G. (1979). Scaling Concepts in Polymer Physics (Ithaca, NY: Cornell University Press).

Ding, F., Dokholyan, N.V., Buldyrev, S.V., Stanley, H.E., and Shakhnovich, E.I. (2002). Direct molecular dynamics observation of protein folding transition state ensemble. Biophys. J. *83*, 3525–3532.

Doi, M., and Edwards, S.F. (1988). The Theory of Polymer Dynamics (New York: Oxford University Press).

Dokholyan, N.V., and Shakhnovich, E.I. (2001). Understanding hierarchical protein evolution from first principles. J. Mol. Biol. *312*, 289–307.

Dokholyan, N.V., Buldyrev, S.V., Stanley, H.E., and Shakhnovich, E.I. (1998). Discrete molecular dynamics studies of the folding of a protein-like model. Fold. Des. *3*, 577–587.

Dokholyan, N.V., Buldyrev, S.V., Stanley, H.E., and Shakhnovich, E.I. (2000). Identifying the protein folding nucleus using molecular dynamics. J. Mol. Biol. *296*, 1183–1188.

Dokholyan, N.V., Borreguero, J.M., Buldyrev, S.V., Ding, F., Stanley, H.E., and Shakhnovich, E.I. (2003). Identifying importance of amino acids for protein folding from crystal structures. Methods Enzymol. *374*, 616–638.

Fitzkee, N.C., and Rose, G.D. (2004a). Reassessing random-coil statistics in unfolded proteins. Proc. Natl. Acad. Sci. USA *101*, 12497–12502.

Fitzkee, N.C., and Rose, G.D. (2004b). Steric restrictions in protein folding: an alpha-helix cannot be followed by a contiguous beta-strand. Protein Sci. *13*, 633–639.

Flory, P.J. (1965). Principles of Polymer Chemistry (Ithaca, NY: Cornell University Press).

Go, N. (1983). Theoretical studies of protein folding. Annu. Rev. Biophys. Bioeng. *12*, 183–210.

Goldenberg, D.P. (2003). Computational simulation of the statistical properties of unfolded proteins. J. Mol. Biol. *326*, 1615–1633.

Kohn, J.E., Millett, I.S., Jacob, J., Zagrovic, B., Dillon, T.M., Cingel, N., Dothager, R.S., Seifert, S., Thiyagarajan, P., Sosnick, T.R., et al. (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. Proc. Natl. Acad. Sci. USA *101*, 12491–12496.

Le Guillou, J.C., and Zinn-Justin, J. (1977). Critical exponents for the n-vector model in three dimensions from field theory. Phys. Rev. Lett. *39*, 95–98.

Levinthal, C. (1968). Are there pathways for protein folding? J. Chem. Phys. *65*, 44–45.

Pappu, R.V., Srinivasan, R., and Rose, G.D. (2000). The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. Proc. Natl. Acad. Sci. USA *97*, 12565–12570.

Shortle, D., and Ackerman, M.S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. Science *293*, 487–489.

Srinivasan, R., and Rose, G.D. (2002). Methinks it is like a folding curve. Biophys. Chem. *101–102*, 167–171.

Tanford, C., Kawahara, K., Lapanje, S., Hooker, T.M., Jr., Zarlengo, M.H., Salahuddin, A., Aune, K.C., and Takagi, T. (1967). Proteins as random coils. 3. Optical rotatory dispersion in 6 M guanidine hydrochloride. J. Am. Chem. Soc. *89*, 5023–5029.

Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A., and Smith, L.J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry *38*, 16424–16431.

Yang, S., Cho, S.S., Levy, Y., Cheung, M.S., Levine, H., Wolynes, P.G., and Onuchic, J.N. (2004). Domain swapping is a consequence of minimal frustration. Proc. Natl. Acad. Sci. USA *101*, 13786–13791.

Yi, Q., Scalley-Kim, M.L., Alm, E.J., and Baker, D. (2000). NMR characterization of residual structure in the denatured state of protein L. J. Mol. Biol. *299*, 1341–1351.

Zhou, Y.Q., and Karplus, M. (1997). Folding thermodynamics of a model three-helix-bundle protein. Proc. Natl. Acad. Sci. USA *94*, 14429–14432.

Zhou, Y.Q., and Karplus, M. (1999). Folding of a model three-helix bundle protein: a thermodynamic and kinetic analysis. J. Mol. Biol. *293*, 917–951.

Zwanzig, R., Szabo, A., and Bagchi, B. (1992). Levinthal's paradox. Proc. Natl. Acad. Sci. USA *89*, 20–22.